

INFINITE ANTI - UNIFORM SOURCES

Daniela G. Tarniceriu¹, Valeriu B. Munteanu¹, Gheorghe Zaharia²

¹“Gheorghe Asachi” Technical University, Faculty of Electronics, Telecommunications and Information Technology, Department of Telecommunications, Bd. Carol I, 700506, Iasi, Romania

²IETR – INSA, UMR CNRS 6164 Rennes, France,

emails: vmuntean@etti.tuiasi.ro, tarniced@etti.tuiasi.ro, gheorghe.zaharia@insa-rennes.fr

Abstract— In this paper we consider the class of anti-uniform Huffman (AUH) codes for sources with infinite alphabet. Poisson, negative binomial, geometric and exponential distributions lead to infinite anti – uniform sources for some ranges of their parameters. Huffman coding of these sources results in AUH codes. We prove that as a result of this encoding, we obtain sources with memory. For these sources we attach the graph and derive the transition matrix between states, the state probabilities and the entropy. If c_0 and c_1 denote the costs for storing or transmission of symbols “0” and “1”, respectively, we compute the average cost for these AUH codes.

Keywords: Huffman coding, average codeword length, code entropy, average cost.

1. Introduction

Consider a discrete source with infinite size alphabet $\xi : (s_1 \ s_2 \ \cdots \ s_k \ \cdots)$ and associated ordered probability distribution $P_n : (p_1 \ p_2 \ \cdots \ p_k \ \cdots)$, where $p_1 \geq p_2 \geq \dots \geq p_k \geq \dots$. It is well known that the Huffman encoding algorithm [1] provides an optimal prefix-free code for this source. A binary Huffman code is usually represented using a binary tree T , whose leaves correspond to the source messages. The two edges emanating from each intermediate tree node (father) are labelled either 0 or 1. The length between the root and a leaf is the length of the binary codeword associated with the corresponding message.

Assuming that $v_k, k = 1, 2, \dots$, is the codeword representing the message s_k , we denote the length of v_k by l_k . The optimality of Huffman coding implies that $l_k \leq l_j$, if $p_k > p_j$.

Anti uniform Huffman (AUH) codes representing a infinite source ξ_n were firstly introduced in [2] and they are characterized by the fact that $l_k = k + 1$, for $k = 0, 1, 2, \dots$.

For this, the following condition has to be fulfilled [2]:

$$\sum_{k=i+2}^{\infty} p_k \leq p_i, 1 \leq i \leq n-3 \quad (1)$$

The AUH codes have been extensively analysed, concerning bounds on average codeword length, entropy and redundancy for different types of probability distribution. In [3] it has been shown that these codes maximize the average length and entropy and tight lower and upper bounds on average codeword length, entropy and redundancy of finite and infinite AUH codes in terms of alphabet size are derived. Related topics are addressed in [4]-[6]. The problem of D-ary Huffman codes is analysed in [7], [8] and it is shown that for AUH codes, by a proper chose of the source probabilities, the average codeword length can be made closer to unity. The problem of bounding the average length of an optimal (Huffman) source code when only limited knowledge of the source symbol probability distribution is available is considered in [9].

The AUH sources appear in a wide variety of situations in the real world, because this class of sources have the property of achieving minimum redundancy in different situations and minimal average cost in highly unbalanced cost regime [10]-[12]. These properties determine a wide range of applications and motivate us to study these sources from an information theoretic perspective. Unequal cost letter problem modeling situations in which different characters have different transmission times or storage costs was addressed in [13], [14]. One example is the telegraph channel with the alphabet $\{., -\}$ in which dashes are twice as long as dots [15]. Another example is the $\{a, b\}$ run – length – limited codes used in magnetic and optical storage, in which the binary codewords are constrained so that each 1 must be preceded by at least a , and at most b , 0's [16]. In [17] the authors have introduced a class of binary prefix codes having the property that each codeword ends in a one. In [18] an application of “1” – ended prefix codes is considered, where it is shown how to construct self-synchronizing codes and how to use them for group testing. A study

concerning the bounds on average codeword length for these codes is performed in [19]. As another example, binary codes whose codewords contain at most a specified number of 1's are used for energy minimization of transmissions in mobile environments [20]. There is a large literature addressing the problem of cost for prefix-free codes with unequal letter cost encoding alphabet [21] and references therein.

AUH sources can be generated by several probability distributions. It has been shown that Poisson, negative binomial, geometric, and exponential distributions lay in the class of AUH sources for some regimes of their parameters [2], [6], [22], [23]. Related topic was addressed in [24], where the authors studied weakly super increasing (WSI) and partial WSI sources in connection with Fibonacci numbers and golden mean, which appeared extensively in modern science and, in particular, have applications in coding and information theory.

The rest of the paper is organized as follows. In Section 2 we present the Huffman encoding of an anti-uniform source with infinite alphabet and show that by Huffman binary encoding a source with memory results. For this source we compute the code entropy $H(X)$. The average cost of the code is also derived. Sections 3, 4, 5 and 6 apply these results for infinite sources with Poisson, negative binomial, geometric and exponential distributions, respectively. We conclude the paper in Section 7.

2. The entropy and the average cost of AUH codes for sources with infinite alphabet

Let there be a discrete source with infinite alphabet, characterized by the distribution:

$$\xi : \begin{pmatrix} s_0^{(t)} & s_1^{(t)} & s_2^{(t)} & \cdots & s_k^{(t)} & \cdots \\ p_0^{(t)} & p_1^{(t)} & p_2^{(t)} & \cdots & p_k^{(t)} & \cdots \end{pmatrix}, \quad (2)$$

We assume that the source is complete, that is

$$\sum_{k=0}^{\infty} p_k^{(t)} = 1 \quad (3)$$

and relation (1) is fulfilled.

After a binary Huffman encoding of this source, the graph in Fig. 1 results, that is, an infinite anti – uniform code. The terminal nodes $s_k^{(i)}$ represent the messages resulting by binary Huffman encoding and $p_k^{(i)}$ denote their probabilities.

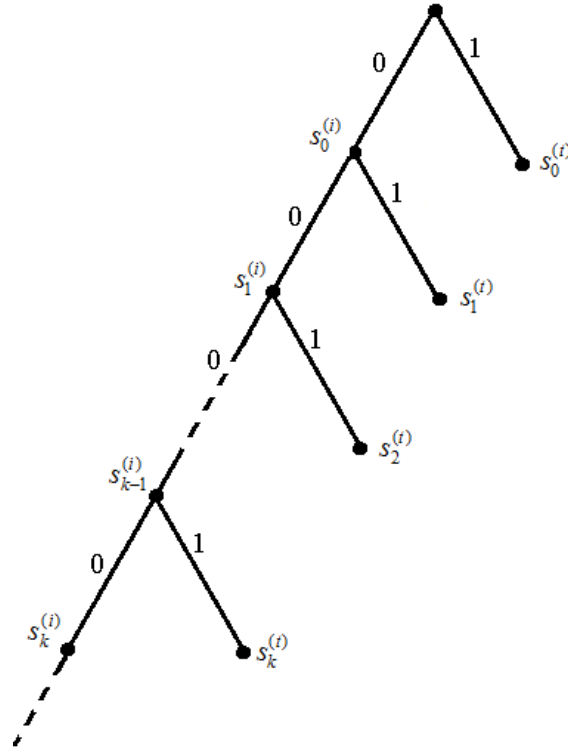


Fig. 1. The graph of Huffman encoding for the source ξ with distribution in (2)

We denote by $s_k^{(i)}$ and $p_k^{(i)}$ the intermediate node “ k ” in the encoding graph and its probability, respectively. Unlike a leaf, an intermediate node is not corresponding to a source message, therefore no probability mass is associated. However, with slight abuse we can call the weight of the intermediate node also probability.

Considering (3), the probabilities of intermediate nodes $p_k^{(i)}$ are obtained recursively, as the sum of the two siblings. In this way, we get:

$$p_k^{(i)} = 1 - \sum_{j=0}^k p_j^{(i)}; k = 0, 1, 2, \dots \quad (4)$$

The structure of the codewords resulting by Huffman binary encoding is:

$$\left. \begin{array}{l} s_0 \rightarrow v_0 \rightarrow 1 \\ s_1 \rightarrow v_1 \rightarrow 01 \\ s_2 \rightarrow v_2 \rightarrow 001 \\ \dots\dots\dots \\ s_k \rightarrow v_k \rightarrow \underbrace{00\dots0}_k 1 \\ \dots\dots\dots \end{array} \right\} \quad (5)$$

The length l_k of the codeword associated with the message $s_k^{(t)}$ is the number of edges on the path between the root and the node $s_k^{(t)}$ in the Huffman tree.

$$l_k = k + 1, \quad k = 0, 1, 2, \dots \quad (6)$$

The average codeword length is determined with

$$\bar{l} = \sum_{k=0}^{\infty} p_k^{(t)} l_k. \quad (7)$$

The branches between successive nodes have the probabilities equal to the ratio between the probability of the node in which the branch ends and the probability of the node from which it starts.

We note that the probabilities to deliver the symbols $x_1 = 1$ or $x_0 = 0$ depend on the node from which they are generated. In other words, as a result of Huffman encoding of the source, a new source $X = \{x_0, x_1\}$ with memory is obtained. Its states correspond to terminal or intermediate nodes (excepting the root) in the graph in Fig. 1. When a terminal node is reached, the binary encoding Huffman procedure is resumed from the graph root.

The graph attached to the source with memory X can be obtained from the Huffman encoding graph of the source ξ , as follows:

- a) We link the terminal nodes in the graph of the source ξ with $S_0^{(t)}$ and $S_0^{(i)}$ through the graph root;
- b) The branch probabilities are the same as in the Huffman encoding graph;

- c) Each terminal or intermediate node (excepting the graph root) will represent a state $S_k^{(t)}$ or $S_k^{(i)}$, $k = 0, 1, 2, \dots$ (as represented in Fig. 2).

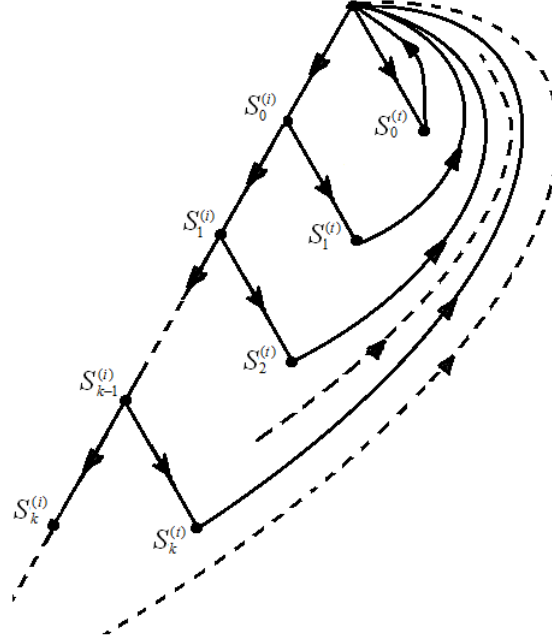


Fig. 2 The graph of the source with memory

Let $S = \{S_0^{(t)}, S_1^{(t)}, \dots, S_{k-1}^{(t)}, S_k^{(t)}, \dots, S_0^{(i)}, S_1^{(i)}, \dots, S_{k-1}^{(i)}, S_k^{(i)}, \dots\}$ be the state set of the source with memory. The probabilities of delivering the symbols $x_0 = 0$ or $x_1 = 1$ from the state $S_{k-1}^{(i)}$, $k = 1, 2, \dots$ are equal to the probabilities of transition from the state $S_{k-1}^{(i)}$, $k = 1, 2, \dots$ to the states $S_k^{(t)}$, $k = 1, 2, \dots$ and $S_k^{(i)}$, $k = 1, 2, \dots$, respectively, i.e.

$$p(x_0 | S_{k-1}^{(i)}) = \frac{p_k^{(i)}}{p_{k-1}^{(i)}}, \quad k = 1, 2, \dots \quad (8)$$

and

$$p(x_1 | S_{k-1}^{(i)}) = \frac{p_k^{(t)}}{p_{k-1}^{(i)}}, \quad k = 1, 2, \dots \quad (9)$$

The transition matrix between states is:

$$\mathbf{T} = \begin{bmatrix} S_0^{(t)} & S_1^{(t)} & S_2^{(t)} & \dots & S_k^{(t)} & \dots & S_0^{(i)} & S_1^{(i)} & \dots & S_k^{(i)} & \dots \\ p_0^{(t)} & 0 & 0 & \dots & 0 & \dots & 1-p_0^{(t)} & 0 & \dots & 0 & \dots \\ p_0^{(t)} & 0 & 0 & \dots & 0 & \dots & 1-p_0^{(t)} & 0 & \dots & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_0^{(t)} & 0 & 0 & \dots & 0 & \dots & 1-p_0^{(t)} & 0 & \dots & 0 & \dots \\ \vdots & \vdots & \vdots & \dots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots \\ 0 & \frac{p_1^{(t)}}{p_0^{(t)}} & 0 & \dots & 0 & \dots & 0 & \frac{p_1^{(i)}}{p_0^{(i)}} & \dots & 0 & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{p_k^{(t)}}{p_{k-1}^{(i)}} & \dots & 0 & 0 & \dots & \frac{p_k^{(i)}}{p_{k-1}^{(i)}} & \dots \\ \vdots & \vdots & \vdots & \dots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots \end{bmatrix} \begin{matrix} S_0^{(t)} \\ S_1^{(t)} \\ \vdots \\ S_k^{(t)} \\ \vdots \\ S_0^{(i)} \\ \vdots \\ S_{k-1}^{(i)} \\ \vdots \end{matrix} \quad (12)$$

Let $\pi_k^{(t)}$ and $\pi_k^{(i)}$, $k=0,1,2,\dots$, denote the state probabilities of the source with memory. They can be determined by means of [25]:

$$[\pi_0^{(t)} \pi_1^{(t)} \dots \pi_k^{(t)} \dots \pi_0^{(i)} \pi_1^{(i)} \dots \pi_k^{(i)} \dots] = [\pi_0^{(t)} \pi_1^{(t)} \dots \pi_k^{(t)} \dots \pi_0^{(i)} \pi_1^{(i)} \dots \pi_k^{(i)} \dots] \mathbf{T} \quad (13)$$

$$\sum_{k=0}^{\infty} (\pi_k^{(t)} + \pi_k^{(i)}) = 1 \quad (14)$$

Considering (7) and (12), from (13) and (14) we get the state probabilities as:

$$\pi_k^{(t)} = \frac{1}{l} p_k^{(t)}, k=0,1,2,\dots \quad (15)$$

$$\pi_k^{(i)} = \frac{1}{l} p_k^{(i)} = \frac{1}{l} \left(1 - \sum_{j=0}^k p_j^{(t)} \right) \quad (16)$$

Generally, the entropy of the source with memory is computed by [26]

$$H(X) = - \sum_{k=0}^{\infty} \sum_{j=0}^1 \pi_k^{(t)} p(x_j | S_k^{(t)}) \log p(x_j | S_k^{(t)}) - \sum_{k=0}^{\infty} \sum_{j=0}^1 \pi_k^{(i)} p(x_j | S_k^{(i)}) \log p(x_j | S_k^{(i)}) \quad (17)$$

Let c_0 and c_1 be the costs associated to the bits 0 and 1, respectively. The average cost of a code is defined by [4]

$$\bar{C} = \sum_{k=0}^{\infty} p_k^{(t)} (n_0(k)c_0 + n_1(k)c_1), \quad (18)$$

where we denote by $n_0(k)$ and $n_1(k)$ the number of 0's and 1's in the codeword corresponding to the source symbol $s_k^{(t)}$.

Considering (5) and (6), the average cost is

$$\bar{C} = \sum_{k=0}^{\infty} p_k^{(t)} (kc_0 + c_1) \quad (19)$$

3. AUH sources with Poisson distribution

Let there be a discrete source with infinite alphabet, characterized by Poisson distribution:

$$\xi: \begin{pmatrix} s_0^{(t)} & s_1^{(t)} & s_2^{(t)} & \dots & s_k^{(t)} & \dots \\ p_0^{(t)} = e^{-\lambda} & p_1^{(t)} = e^{-\lambda} \frac{\lambda}{1!} & p_2^{(t)} = e^{-\lambda} \frac{\lambda^2}{2!} & \dots & p_k^{(t)} = e^{-\lambda} \frac{\lambda^k}{k!} & \dots \end{pmatrix}, \quad (20)$$

In [22] it is shown that any Poisson distribution with parameter $\lambda \leq 1$ satisfies condition (1) and leads to an AUH code.

The source is complete, because

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1 \quad (21)$$

Theorem 1

The entropy and the average cost of the code resulting by binary encoding of AUH source with Poisson distribution are determined by:

$$H(X) = \frac{1}{l} \left(\lambda \log e - \lambda \log \lambda + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \log(k!) \right) \quad (22)$$

$$\bar{C} = c_0 \lambda + c_1 \quad (23)$$

Proof

Considering (20) and (4), the probabilities of intermediate nodes are obtained by:

$$p_k^{(i)} = 1 - \sum_{j=0}^k e^{-\lambda} \frac{\lambda^j}{j!}; k = 0, 1, 2, \dots \quad (24)$$

The average codeword length is obtained considering (20) and (6) in (7)

$$\bar{l} = \lambda + 1. \quad (25)$$

The probabilities of delivering the symbols $x_1 = 1$ or $x_0 = 0$ from the state $S_{k-1}^{(i)}$, $k = 1, 2, \dots$ are

$$p(x_1 | S_{k-1}^{(i)}) = \frac{p_k^{(i)}}{p_{k-1}^{(i)}} = \frac{e^{-\lambda} \frac{\lambda^k}{k!}}{1 - \sum_{j=0}^{k-1} e^{-\lambda} \frac{\lambda^j}{j!}}, k = 1, 2, \dots \quad (26)$$

and

$$p(x_0 | S_{k-1}^{(i)}) = \frac{p_k^{(i)}}{p_{k-1}^{(i)}} = \frac{1 - \sum_{j=0}^k e^{-\lambda} \frac{\lambda^j}{j!}}{1 - \sum_{j=0}^{k-1} e^{-\lambda} \frac{\lambda^j}{j!}}, k = 1, 2, \dots \quad (27)$$

respectively.

The probabilities of delivering the symbols $x_1 = 1$ or $x_0 = 0$ from the state $S_k^{(t)}$, $k = 0, 1, 2, \dots$ are

$$p(x_1 | S_k^{(t)}) = e^{-\lambda}, k = 0, 1, 2, \dots \quad (28)$$

and

$$p(x_0 | S_k^{(t)}) = 1 - e^{-\lambda}, k = 0, 1, 2, \dots \quad (29)$$

respectively.

Considering (20), (24) and (25) into (15) and (16), we get the stationary state probabilities:

$$\pi_k^{(t)} = \frac{1}{l} \left(e^{-\lambda} \frac{\lambda^k}{k!} \right), k = 0, 1, 2, \dots \quad (30)$$

$$\pi_k^{(i)} = \frac{1}{l} \left(1 - \sum_{j=0}^k e^{-\lambda} \frac{\lambda^j}{j!} \right), k = 0, 1, 2, \dots \quad (31)$$

Substituting (26) - (31) into (17), we get the entropy of the source with memory in (22).

The average cost of the AUH code for the source with Poisson distribution, given in (23), is obtained by substituting (20) into (19).

4. AUH sources with negative binomial distribution

Let there be a discrete source with infinite alphabet, characterized by the negative binomial distribution:

$$\xi : \begin{pmatrix} S_0^{(t)} & S_1^{(t)} & S_2^{(t)} & \dots & S_k^{(t)} & \dots \\ p_0^{(t)} = C_{r-1}^{r-1} p^r & p_1^{(t)} = C_r^{r-1} p^r q & p_2^{(t)} = C_{r+1}^{r-1} p^r q^2 & \dots & p_k^{(t)} = C_{r+k-1}^{r-1} p^r q^k & \dots \end{pmatrix}, \quad (32)$$

where $q=1-p$.

The source is complete, because

$$\sum_{k=0}^{\infty} C_{r+k-1}^{r-1} p^r q^k = 1 \quad (33)$$

Theorem 2

The entropy and the average cost of the code resulting by binary encoding of AUH source with negative binomial distribution are determined by:

$$H(X) = -\frac{1}{\bar{l}} \left(r \log p + \frac{qr}{p} \log q + p^r \sum_{k=1}^{\infty} C_{r+k-1}^{r-1} q^k \log C_{r+k-1}^{r-1} \right) \quad (34)$$

$$\bar{C} = c_0 \frac{qr}{p} + c_1 \quad (35)$$

Proof

Considering (32) and (4), the probabilities of intermediate nodes are obtained by:

$$p_k^{(i)} = 1 - \sum_{j=0}^k C_{r+j-1}^{r-1} p^r q^j; \quad k = 0, 1, 2, \dots \quad (36)$$

The average codeword length is obtained considering (32) and (6) in relation (7)

$$\bar{l} = \sum_{k=0}^{\infty} (k+1) C_{r+k-1}^{r-1} p^r q^k = \frac{rq + p}{p}. \quad (37)$$

The probabilities of delivering the symbols $x_1 = 1$ or $x_0 = 0$ from the state $S_{k-1}^{(i)}$, $k = 1, 2, \dots$ are

$$p(x_1 | S_{k-1}^{(i)}) = \frac{p_k^{(i)}}{p_{k-1}^{(i)}} = \frac{C_{r+k-1}^{r-1} p^r q^k}{1 - \sum_{j=0}^{k-1} C_{r+j-1}^{r-1} p^r q^j}, \quad k = 1, 2, \dots \quad (38)$$

and

$$p(x_0 | S_{k-1}^{(i)}) = \frac{p_k^{(i)}}{p_{k-1}^{(i)}} = \frac{1 - \sum_{j=0}^k C_{r+j-1}^{r-1} p^r q^j}{1 - \sum_{j=0}^{k-1} C_{r+j-1}^{r-1} p^r q^j}, \quad k = 1, 2, \dots \quad (39)$$

respectively.

The probabilities of delivering the symbols $x_1 = 1$ or $x_0 = 0$ from the state $S_k^{(t)}$, $k = 0, 1, 2, \dots$ are

$$p(x_1 | S_k^{(t)}) = C_{r-1}^{r-1} p^r = p^r, k = 0, 1, 2, \dots \quad (40)$$

and

$$p(x_0 | S_k^{(t)}) = 1 - C_{r+k-1}^{r-1} p^r = 1 - p^r, k = 0, 1, 2, \dots \quad (41)$$

respectively.

Considering (32), (36) and (37) in (15) and (16), we get the stationary state probabilities:

$$\pi_k^{(t)} = \frac{1}{l} (C_{r+k-1}^{r-1} p^r q^k), k = 0, 1, 2, \dots \quad (42)$$

$$\pi_k^{(i)} = \frac{1}{l} \left(1 - \sum_{j=0}^k C_{r+j-1}^{r-1} p^r q^j \right), k = 0, 1, 2, \dots \quad (43)$$

Substituting (38) - (43) into (17), we get the entropy of the source with memory, given in (34).

We obtain the average cost of the AUH code in (35) for the source with negative binomial distribution, by substituting (32) in (19).

5. AUH sources with geometric distribution

Let there be a discrete source with infinite alphabet, characterized by the geometric distribution:

$$\xi : \begin{pmatrix} s_0^{(t)} & s_1^{(t)} & s_2^{(t)} & \dots & s_k^{(t)} & \dots \\ p_0^{(t)} = q & p_1^{(t)} = pq & p_2^{(t)} = p^2 q & \dots & p_k^{(t)} = p^k q & \dots \end{pmatrix}, \quad (44)$$

where $q=1-p$.

In [23] it is shown that geometric distribution with parameter $0 < p \leq (\sqrt{5}-1)/2$ satisfies condition (1) and leads to an AUH code.

The source is complete, because

$$\sum_{k=0}^{\infty} p^k q = 1 \quad (45)$$

Theorem 3

The entropy and the average cost of the code resulting by binary encoding of AUH source with geometric distribution are determined by:

$$H(X) = -\frac{1}{l} \left(\log(1-p) + \frac{p}{1-p} \log p \right) \quad (46)$$

$$\bar{C} = c_0 \frac{p}{1-p} + c_1 \quad (47)$$

Proof

Considering (44) and (4), the probabilities of intermediate nodes are obtained by:

$$p_k^{(i)} = p^{k+1}; k = 0, 1, 2, \dots \quad (48)$$

The average codeword length is obtained considering (44) and (6) in (7)

$$\bar{l} = \sum_{k=0}^{\infty} (k+1) q p^k = \frac{1}{1-p}. \quad (49)$$

The probabilities of delivering the symbols $x_1 = 1$ or $x_0 = 0$ from the state $S_{k-1}^{(i)}$, $k = 1, 2, \dots$ are

$$p(x_1 | S_{k-1}^{(i)}) = q = 1-p, k = 1, 2, \dots \quad (50)$$

and

$$p(x_0 | S_{k-1}^{(i)}) = p, k = 1, 2, \dots \quad (51)$$

respectively.

The probabilities of delivering the symbols $x_1 = 1$ or $x_0 = 0$ from the state $S_k^{(i)}$, $k = 0, 1, 2, \dots$ are

$$p(x_1 | S_k^{(i)}) = 1-p, k = 0, 1, 2, \dots \quad (52)$$

and

$$p(x_0 | S_k^{(i)}) = p, k = 0, 1, 2, \dots \quad (53)$$

respectively.

Considering (44), (48) and (49) in (15) and (16), we get the stationary state probabilities:

$$\pi_k^{(i)} = \frac{1}{l} (1-p) p^k \quad (54)$$

$$\pi_k^{(i)} = \frac{1}{l} p^{k+1} \quad (55)$$

Substituting (50) - (55) in (17), we get the entropy of the source with memory, given in (46).

We obtain the average cost of the AUH code in (47) for the source with geometric distribution substituting (44) in (19).

6. AUH sources with exponential distribution

Let there be a discrete source with infinite alphabet, characterized by the exponential distribution:

$$\xi : \begin{pmatrix} s_0^{(t)} & s_1^{(t)} & s_2^{(t)} & \cdots & s_k^{(t)} & \cdots \\ p_0^{(t)} = 1 - e^{-\lambda} & p_1^{(t)} = e^{-\lambda}(1 - e^{-\lambda}) & p_2^{(t)} = e^{-2\lambda}(1 - e^{-\lambda}) & \cdots & p_k^{(t)} = e^{-k\lambda}(1 - e^{-\lambda}) & \cdots \end{pmatrix}, \quad (56)$$

In [6] it is shown that the discrete form of exponential distribution is anti – uniform iff $\lambda \geq \ln\{(1 + \sqrt{5}) / 2\} \approx 0.4812$.

The source is complete, because

$$\sum_{k=0}^{\infty} e^{-k\lambda}(1 - e^{-\lambda}) = 1 \quad (57)$$

Theorem 4

The entropy of the source ξ , as well as the entropy and the average cost of the source with memory resulted by binary encoding of AUH source with exponential distribution are determined by:

$$H(X) = -\frac{1}{l} \left(\log(1 - e^{-\lambda}) + \frac{e^{-\lambda} \log e^{-\lambda}}{1 - e^{-\lambda}} \right) \quad (58)$$

$$\bar{C} = c_0 \frac{e^{-\lambda}}{1 - e^{-\lambda}} + c_1 \quad (59)$$

Proof

Considering (56) and (4), the probabilities of intermediate nodes are obtained by:

$$p_k^{(i)} = e^{-(k+1)\lambda}, k = 0, 1, 2, \dots \quad (60)$$

The average codeword length is obtained considering (56) and (6) in relation (7)

$$\bar{l} = \sum_{k=0}^{\infty} (k+1) e^{-k\lambda} (1 - e^{-\lambda}) = \frac{1}{1 - e^{-\lambda}} \quad (61)$$

The probabilities of delivering the symbols $x_1 = 1$ or $x_0 = 0$ from the state $S_{k-1}^{(i)}$, $k = 1, 2, \dots$ are

$$p(x_1 | S_{k-1}^{(i)}) = 1 - e^{-\lambda}, k = 1, 2, \dots \quad (62)$$

and

$$p(x_0 | S_{k-1}^{(i)}) = e^{-\lambda}, k = 1, 2, \dots \quad (63)$$

respectively.

The probabilities of delivering the symbols $x_1 = 1$ or $x_0 = 0$ from the state $S_k^{(t)}$, $k = 0, 1, 2, \dots$ are

$$p(x_1 | S_k^{(t)}) = 1 - e^{-\lambda}, k = 0, 1, 2, \dots \quad (64)$$

and

$$p(x_0 | S_k^{(t)}) = e^{-\lambda}, k = 0, 1, 2, \dots \quad (65)$$

respectively.

Considering (56), (60) and (61) in (15) and (16), we get the stationary state probabilities as:

$$\pi_k^{(t)} = \frac{1}{l} (1 - e^{-\lambda}) e^{-k\lambda} \quad (66)$$

$$\pi_k^{(i)} = \frac{1}{l} e^{-(k+1)\lambda} \quad (67)$$

Substituting (62) - (67) in (17), we get the entropy of the source with memory, given in (58).

We obtain the average cost of the AUH code in (59) for the source with exponential distribution, substituting (56) in (19).

7. Conclusions

In this paper we have considered the class of AUH sources with infinite alphabets. We have shown that performing a binary Huffman encoding of these sources, sources with memory result. For these sources we have build the encoding graph and have specified the rules for drawing the graph of the source with memory $X = \{x_0 = 0, x_1 = 1\}$. The graph of the source with memory is obtained from the encoding graph by linking the terminal nodes with the states $S_0^{(t)}$ and $S_0^{(i)}$ through the graph root. The states of the source with memory correspond to the terminal or intermediate nodes in the encoding graph, excepting the root. We have determined in the general case the state probabilities of the source with memory, as well as the transition probabilities between states. The

entropy of this source with memory is computed. We assumed the costs c_0 and c_1 for the symbols “0” and “1”, respectively, and compute the average cost for these Huffman codes. Obviously, the Huffman encoding procedure assures minimum average length, but the average cost is not minimum. It can be easily verified that if the costs of symbols “0” and “1” are equal to unity, the average cost becomes equal to the average length. We applied the results for several AUH sources with Poisson, negative binomial, geometric and exponential distributions.

8. References

- [1] Huffman R. A method for the construction of minimum – redundancy codes. Proc. IRE 1952; 40: p. 1098 – 1101.
- [2] Esmaeili M, Kakhbod A. On antiuniform and partially antiuniform sources. Proc. IEEE ICC, 2006. p. 1611 – 5.
- [3] Mohajer S, Kakhbod A. Anti – uniform Huffman codes. IET Commun. 2011; 5: p. 1213 – 9.
- [4] Mohajer S, Kakhbod A. Tight bounds on the AUH codes. Proceedings on Information Sciences and Systems, Princeton, New Jersey, USA, 2008. p. 1010 – 4.
- [5] Mohajer S, Pakzad P, Kakhbod A. Tight bounds on the redundancy of Huffman codes. Proceedings on IEEE ITW, Punta del Este 2006. p. 131 – 5.
- [6] Esmaeili M, Kakhbod A. On information theory parameters of infinite anti-uniform sources. IET Commun 2007; 1: p. 1039 – 1041.
- [7] Zaharia G, Munteanu V, Tarniceriu D. Tight bounds on the codeword length and average codeword length for D-ary Huffman codes – Part 1. Proceedings IEEE ISSCS, Iasi, Romania 2009. p. 531-4.
- [8] Zaharia G, Munteanu V, Tarniceriu D. Tight bounds on the codeword length and average codeword length for D-ary Huffman codes – Part 2. Proceedings IEEE ISSCS, Iasi, Romania 2009. p. 535-8.
- [9] Cicalese F, Vaccaro U. Bounding the Average Length of Optimal Source Codes via Majorization Theory. Trans. Inf. Theory 2004; 50: p. 633-637.
- [10] Johnsen O. On the redundancy of binary Huffman codes. IEEE Trans. Inf. Theory 1980; p. 26:220-222.
- [11] Bradford P, Golin M, Larmore LL and Rytter W. Optimal prefix free codes for unequal letter costs and dynamic programming with the Monge property. J. Algorithms, 2002; 42: p. 277 – 303.
- [12] Gilbert E N. Coding with digits of unequal costs. Trans. Inf. Theory 1995; 41: p. 596-600.
- [13] Blachman NM. Minimum cost coding of information. IRE Trans. Inf. Theory 1954; PGIT-3:1 p. 39-149.

- [14] Varn B. Optimal variable length codes (arbitrary symbol cost and equal code word probability). *Inf. Contr.* 1971; 19: p. 289-301.
- [15] Krause R M. Channels which transmit letters of unequal duration. *Inf. Contr.* 1962; 55: p. 13 – 24.
- [16] Golin M, Rote G. A dynamic programming algorithm for constructing optimal prefix-free codes for unequal costs. *IEEE Trans. Inf. Theory* 1998; 44(5): p. 1770-1781.
- [17] Berger T, Yeung R. Optimum 1-ended binary prefix codes. *Trans. Inf. Theory* 1990; 36(6): p. 1434-1441.
- [18] Capocelli RM, De Santis A, Gargano L, Vaccaro U. On the construction of statistically synchronizable codes. *IEEE Trans. Inf. Theory* 1992; 38: p. 407-414.
- [19] Capocelli RM, De Santis A, Persiano G. Binary prefix codes ending in a 1 *Trans. Inf. Theory* 1994; 40: p. 1296-1302.
- [20] Korach E, Dolov S, Yukelson D. The sound of silence: Guessing games for saving energy in mobile environment. *Proc. Of the Eighteenth Annual Joint Conference of IEEE Computer and Communications Societies (IEEE INFOCOM'99)* 1999. p. 768-75.
- [21] Golin M, Li J. More efficient algorithms and analyses for unequal letter cost prefix-free coding. *Trans. Inf. Theory* 2008; 54(8): p. 3412-3424.
- [22] Humblet P. Optimal source coding for a class of integer alphabets. *IEEE Trans. Inf. Theory* 1978; 24(1): p. 110 – 112.
- [23] Gallager R, Van Voorhis D. Optimal source coding for geometrically distributed integer alphabets. *IEEE Trans. Inf. Theory* 1975; 21(2): p. 228 – 230.
- [24] Esmaeili M, Gulliver A, Kakhbod A. The Golden Mean, Fibonacci Matrices and Partial Weakly Super-Increasing Sources. *Chaos, Solitons and Fractals* 2009; 42(1): p. 435 – 440.
- [25] Kemeny JG, Snell TL. *Finite Markov Chains*. Princeton, NJ: Van Nostrand, 1960.
- [26] Cover TM, Thomas JA. *Elements of Information Theory*. John Wiley and Sons, Inc. New York, 1991.